Published online 21 November 2013 Nucleic Acids Research, 2014, Vol. 42, Database issue **D1007–D1012**
doi:10.1093/nar/gkt1166

Lynx: a database and knowledge extraction engine for integrative medicine

Dinanath Sulakhe[1,*], Sandhya Balasubramanian[2], Bingqing Xie[2,3], Bo Feng[2,3], Andrew Taylor[2], Sheng Wang[4], Eduardo Berrocal[2,3], Utpal Dave[1], Jinbo Xu[4], Daniela Börnigen[2,4], T. Conrad Gilliam[1,2] and Natalia Maltsev[1,2,*]

[1]Computation Institute, the University of Chicago, Chicago, IL 60637, USA, [2]Department of Human Genetics, the University of Chicago, Chicago, IL 60637, USA, [3]Department of Computer Science, Illinois Institute of Technology, Chicago, IL 60616, USA and [4]Toyota Technological Institute at Chicago, Chicago, IL 60637, USA

Received August 23, 2013; Revised October 28, 2013; Accepted October 29, 2013

ABSTRACT

We have developed Lynx (http://lynx.ci.uchicago.edu)—a web-based database and a knowledge extraction engine, supporting annotation and analysis of experimental data and generation of weighted hypotheses on molecular mechanisms contributing to human phenotypes and disorders of interest. Its underlying knowledge base (LynxKB) integrates various classes of information from >35 public databases and private collections, as well as manually curated data from our group and collaborators. Lynx provides advanced search capabilities and a variety of algorithms for enrichment analysis and network-based gene prioritization to assist the user in extracting meaningful knowledge from LynxKB and experimental data, whereas its service-oriented architecture provides public access to LynxKB and its analytical tools via user-friendly web services and interfaces.

INTRODUCTION

Technological advances in genomics now allow us to produce biological data at unprecedented tera- and petabyte scales. The extraction of useful knowledge from these voluminous data sets critically depends on seamless integration of clinical, genomic and experimental information with prior knowledge about genotype–phenotype relationships accumulated in a plethora of databases. Furthermore, these large and complex integrated knowledge bases should be accessible to search engines and algorithms that drive efficient knowledge extraction advancing scientific insight and the development of biomedical applications.

To meet these challenges, we developed Lynx (http://lynx.ci.uchicago.edu), a web-based database and a knowledge extraction engine for annotation and analysis of high-throughput biomedical data. Lynx database was designed specifically to support both discovery-based and hypothesis-based approaches to prediction of genetic factors and networks contributing to phenotypes of interest. Such unique support is provided by integration of vast amounts of information (e.g. genomic data, pathways and molecular interactions and other) from public and private repositories, as well as the targeted acquisition of phenotypic information and data describing association of genetic factors with diseases, clinical symptoms and phenotypic features. Lynx advanced search engines and a variety of algorithms for enrichment analysis and network-based gene prioritization support the extraction of meaningful knowledge from LynxKB and experimental data provided by the users. Lynx also enables formulation of weighted hypotheses regarding molecular mechanisms contributing to human phenotypes and disorders of interest.

LYNX DESIGN AND COMPONENTS

The Lynx database system has the following major components: (i) Integrated Lynx knowledge base (LynxKB); (ii) Knowledge extraction services currently available for LynxKB, including advanced search capabilities, features-based gene enrichment analysis and network-based gene prioritization, which may be invoked via the Lynx REST

*To whom correspondence should be addressed. Tel: +1 773 702-4960; Fax: +1 773 834-0505; Email: maltsev@uchicago.edu
Correspondence may also be addressed to Dinanath Sulakhe. Tel: +1 630 252 7856; Fax: +1 630 252 5676; Email: sulakhe@mcs.anl.gov
Present address:
Natalia Maltsev, Human Genetics Department, the University of Chicago, CLSC, E. 58th str. Chicago, IL, 60637, USA.
Dinanath Sulakhe, Computation Institute, Chicago, IL 60637, USA.

interface; and (iii) 'Web Interface', a user-friendly web interface for accessing the annotations and analytical tools.

Lynx integrated knowledgebase

LynxKB is a database integrating modeled data from >35 databases and manually curated private collections (Table 1). These data are used for annotation and extraction of knowledge from LynxKB via database queries or from experimental data provided by the user. An XML schema-driven annotation service supports annotations from the LynxKB as RESTful web services. Additionally, LynxKB contains a number of manually curated in-house data collections, including *inter alia* customized ontologies for early brain development and brain connectivity (developed in collaboration with Dr. Paciorkowski, University of Rochester), weighted collections of candidate genes provided by our clinical collaborators or extracted from Developmental Brain Disorders Database (DBDB) and other disease-related data sources such as AutDB (19), Schizophrenia Gene Resource (20), LisDB (https://lisdb.ci.uchicago.edu) and Cancer Gene Index (https://wiki.nci.nih.gov/display/cageneindex/caBIO). Lynx also provides an exclusive analytical access to the text-mining data describing molecular interactions from GeneWays (26). Integration of the data describing clusters of transcription factors binding sites (28) and enhancers (29), as provided by the Vista project, allows one to factor the information regarding non-coding genomic signals into the Lynx predictions of genetic factors involved in disorders of interest. Integrated structured data from Lynx KB is available for downloads in multiple formats (e.g. XML, CSV, TXT, JSON) via a web-based user interface and web services.

Table 1. Data types and resources integrated in LynxKB

Type of data	Source
Genomic	NCBI (1), Ensembl (2), UniGene (3), TRANSFAC[b] (4), RefSeq (5)
Proteomic	BIND (6), BioGRID (7), HPRD (8), MINT (9), UniProt (10), InterPro (11)
Pathways-related	KEGG (12), Reactome (13), NCI (14), BioCarta, STRING[b] (15), TRANSPATH[b] (16), Pathway Commons (17)
Disease-specific	OMIM, Disease ontology (18), AutDB (19), SZGR (20), Cancer gene index, AGRE, DBDB[a], LisDB[a]
Phenotypic	OMIM, Human phenotype ontology (21), customized ontologies[a]
Variations	Genomic association database (22), Database of genomic variants (23), Human genomic mutation database[b] (24), SLEP (25), NHGRI
Text-mining	GeneWays[a] (26), Diseases (University of Copenhagen)
Pharmacogenomics	Comparative toxicogenomics database (CTD) (27)

[a]Customized and manually curated sources of information.
[b]The resources are not displayed on the annotations page due to the proprietary license restrictions and/or are used exclusively in the analytical pipelines.

Lynx data are available for download in a number of ways: (i) 'Lynx KB database dumps'. Due to the fact that public data are available for download at the respective sources and the size of a complete integrated Lynx KB is prohibitively large, downloading the full content of Lynx KB may be impractical. However, any part of the public data integrated into Lynx KB is available for download in the form of tab-delimited tables and database dumps on request; (ii) all annotations and results of analysis in Lynx are available for download in CSV format via the 'download' button displayed on every page; and (iii) any Lynx object or set of objects as well as the results of annotation and analysis may be downloaded using web services in JSON and XML format.

Lynx knowledge extraction engine

Seamless integration of data, knowledge-extraction services and integrative analysis in Lynx provide a one-stop solution for generating weighted hypotheses regarding the molecular mechanisms contributing to the phenotypes of interest (Figure 1). Lynx supports multiple entry points for annotation and analysis of individual objects (e.g. genes, pathways, disorders) and batch queries. The user can submit search-based queries to LynxKB or

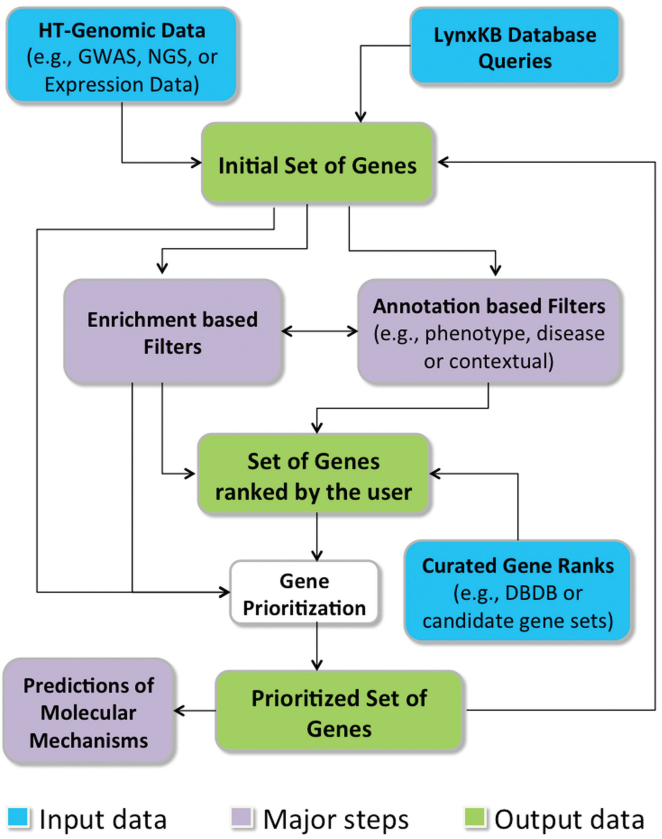

Figure 1. A workflow of knowledge extraction in the Lynx database where initial query genes are filtered interactively using annotations or based on the results of enrichment analysis. Resulting gene sets are ranked by the user according to his/her preferences and further prioritized using networks-based prioritization assisting in the prediction of molecular mechanisms contributing to the phenotype or biological process of interest to the user.

experimental results to be analyzed by Lynx, such as the results of next-generation sequencing (NGS), copy number variation-based analyses or gene expression data in the form of SNPs, genomic coordinates or gene lists to be annotated or downstream analyzed via the web user interface and its integrated services. Lynx provides the following knowledge extraction tools for the downstream analyses and annotations:

Advanced search

The large-scale integration of biomedical data in Lynx provides a great opportunity to mine these data with a systems perspective in mind. Its powerful search capabilities, based on Apache Lucene (http://lucene.apache.org/), allow users to generate highly selective data sets by filtering the queries to LynxKB on multiple parameters of interest to the user (e.g. phenotypes, pathways, keywords), as illustrated below in the case study, as well as highly efficient search functionality based on phrase queries, wildcard queries and Boolean operators for a deeper refinement of search results. Additionally, as illustrated below in the case study, users can start with broader searches based on diseases, pathways or symptoms of interest and then further refine and narrow down the results of the searches according to the parameters of interest. Another important feature of the advanced search functionality in Lynx is that the results of the queries are presented in association with other relevant annotations, such as genes, pathways, tissues, phenotypes and more, to provide a comprehensive overview for an object of interest. Lynx's advanced search capabilities provide a unique perspective on the biological data of interest and can be an extremely powerful tool for researchers.

Annotation services

Lynx's XML schema-driven annotation service provides annotations from the integrated database as RESTful web services. Every query to the LynxKB for an individual object (e.g. a gene) or a batch query (e.g. list of genes or genomic coordinates) extracts all information relevant to the query from LynxKB for the growing list of annotations [e.g. gene function description (RefSeq), associated pathways, diseases, clinical symptoms, molecular interactions, toxicogenomic information, Gene Ontology categories, tissues and other related annotations] and displays it to the user according to his/her preferences. Lynx provides detailed web interfaces for single-gene or multiple-gene annotations that allow users to get a complete understanding of the functionality of the genes of interest from various different perspectives. All information related to the objects is easily accessible via user interface and available for download in tab-delimited, XML or JSON formats (web services).

Statistical enrichment analysis

Lynx assists the user in formulating the hypotheses regarding the molecular mechanisms involved in the phenomena under study by providing tools for enrichment analysis and identification of functional categories over-represented in the query data sets. Two singular enrichment analysis algorithms, Bayes factor and *P*-value estimates are used in our pipeline for this purpose (see Xie *et al.* for more description and results of analysis (30)). Enrichment analysis in Lynx is based on a large variety of features obtained from multiple sources [e.g. associated pathways and diseases (Table 1), various levels of resolution of Gene Ontology terms], as well as unique-for-the-system customized brain development and brain connectivity ontologies, symptoms-level phenotypes and associated non-coding signals (e.g. enhancers and clusters of transcription factors binding sites). The results of the enrichment analyses based on multiple categories of interest to the user may be used for formulating a working hypothesis regarding molecular mechanisms involved in phenomena of interest. Lynx also supports contextual enrichment analysis (e.g. against genes expressed in a particular tissue or on a particular developmental stage) that may substantially increase the accuracy of the results.

Network-based gene prioritization

Gene prioritization proposes promising candidate genes from a large set of genes or even from the entire genome for a disease or phenotype of interest. Here, for network-based gene prioritization, Lynx integrates five network propagation algorithms [simple random walk, heat kernel diffusion (31), PageRank with priors (32), HITS with priors (33) and K-step Markov (33)], and using STRING version 9.0 (15) as the underlying protein interaction network as initially suggested in PINTA (34,35). To use known disease genes as input, the algorithms were accordingly modified for Lynx by replacing the continuous microarray expression data—as requested from the original PINTA implementation—with binary data using seed genes associated with a disease or phenotype of interest: a '1' is fed as an input for each seed gene, whereas a '0' is assigned to all non-seed genes (36). Additionally, these algorithms were modified to accommodate a variety of weighted data types to be used for gene prioritization including ranked gene to phenotype associations, weighted canonical pathways, gene expression, NGS data and others. Consequently, the propagation algorithms for gene prioritization provide a ranked list of novel and promising candidate genes based on the propagated signal through the network, starting from binary data associated with disease related genes in the network.

CASE STUDY: IDENTIFICATION OF MOLECULAR MECHANISMS ASSOCIATED WITH SEIZURES IN AUTISM

This case study will illustrate the functionality of Lynx by predicting genes and molecular mechanisms associated with a particular symptom of autism (seizures) based on various Lynx analyses, such as annotation, gene set enrichment analysis and gene prioritization.

Autism spectrum disorders (ASD) are known to be associated with an increased incidence of epilepsy and of epileptiform discharges on electroencephalograms. However, it is unknown whether epileptiform discharges

correlate with symptoms of ASD and what are the contributing molecular mechanisms (37,38).

To formulate a weighted hypothesis regarding genes and molecular mechanisms potentially contributing to epilepsy in patients with autism, we have performed the following steps:

Step 1: Lynx advanced search was used to perform 'fuzzy' search for autism candidate genes against 'disease' object. The search returned 483 genes associated with autism by OMIM, AutDB and Disease Database from the University of Copenhagen. These genes were further filtered using 'seizures' as a fuzzy search term. The resulting query returned 59 genes positively associated both with autism and seizures phenotype (Supplementary Table S1).

Step 2: The enrichment analysis of these 59 genes associated both with autism and seizures showed over-representation of the functional categories associated with synaptic transmission and ionotropic glutamate receptor binding and voltage-gated sodium channel activity already known to be associated with ASD and epileptic phenotypes.

Step 3: The 59 genes obtained in Step 1 can be ranked according to the strength of their association with autism, as suggested by AutDB or expert curation, or can be assigned a default score of '1' as shown in the use case.

Step 4: The ranked set of genes from Step 3 was used as an input to the gene prioritization tool, based on the heat kernel-ranking algorithm (Supplementary Table S2). Default parameters were used to run the algorithm.

The results of gene prioritization allowed predicting additional 31 high-scoring genes ($P = <0.02$) potentially contributing to epileptic phenotype in autistic patients (Supplementary Table S3). A number of these genes predicted by the network were recently found to be associated with ASD and epileptic phenotypes, but not yet included in AutDB and OMIM databases (and consequently in LynxDB) as markers for ASD. These include: DLG3, discs, large homolog 3 (39), GAD1, glutamate decarboxylase 1, brain type (40), DOCK8, dedicator of cytokinesis 8 (41), GABRB3, GABA A receptor, beta 3 (42), GLUD2, glutamate dehydrogenase 2 (43) and others (see Supplementary Table S3 for more details). All results of analyses are available for download in various formats via user interface or web services. A video and tutorial describing this and other examples of using Lynx for data annotation and analyses are available at the Lynx Web site at http://lynx.ci.uchicago.edu/usecase.html.

SYSTEMS ARCHITECTURE

Lynx is designed using a service-oriented architecture and is implemented using JAX-RS and Spring framework, (44) to provide the integrated data and analytical tools as RESTful services (45). The integrated data are modeled and represented as XML schemas and using JAXB (46) are automatically translated into Java objects that are then used to encapsulate data from the MySQL database. The resulting annotations and results of analysis are delivered in XML, JSON or TXT format as per the request. The project is being developed using the Maven (http://maven.apache.org) multi-module architecture so that various data access objects (DAO) modules; service modules and REST-resource modules are independently implemented and reused where necessary using Spring's dependency injection. The algorithms involved in the analytical steps are implemented using Java and required statistical packages (such as Matlab, which is used in network-based prioritization) and integrated within the project as maven modules. The modular design architecture allows us to maintain 'separation of concerns' within the complete project without introducing any design or architecture-based dependencies.

Data and analytical web services

Although the integrated data and annotations as well as the various analytical tools are presented to the users via web interface, the service-oriented architecture enables other users/groups to leverage our work and integrate it within their own research tools and platforms. For example, there are current ongoing efforts by the Globus Genomics project (47) at the University of Chicago Computation Institute to integrate the Lynx Knowledge base annotation services and analytical workflows (via web services) for analysis and annotation of the results of the NGS. The Developmental Brain Disorders Database (https://www.dbdb.urmc.rochester.edu/home) at the University of Rochester and RViewer (48) are also using Lynx RESTful web service interface for annotation of genomic data. End users can download the data sets of interest and results of analysis from the web interface.

CONCLUSIONS

We present the Lynx database and knowledge extraction suite of tools designed specifically to support the discovery and hypothesis-based approaches to identification of genetic factors contributing to phenotypes or disorders of interest. Lynx integrates the main downstream analyses, such as gene annotation, gene set enrichment analysis and gene prioritization within one engine, based on a large knowledge base from public and private data and a powerful search engine that enables the user to access the knowledge base in a user-friendly web interface.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Andrey Rzhetsky, Inna Dubchak, William Dobyns, Kathleen Millen, Alex Paciorkowski and Chalam Chitturi for their invaluable contributions to the project. They acknowledge support

from the Autism Genetic Resource Exchange (AGRE) and Autism Speaks. The authors gratefully acknowledge the resources provided by the AGRE consortium* and the participating AGRE families. The Autism Genetic Resource Exchange (AGRE) is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to Clara M. Lajonchere (PI).

FUNDING

Mr. and Mrs. Lawrence Hilibrand, the Boler Family Foundation and National Institutes of Health/National Institute of Neurological Disorders and Stroke [grant NS050375]; and the Genetic Basis of Mid-Hindbrain Malformations. Funding for open access charge: National Institutes of Health/National Institute of Neurological Disorders and Stroke [grant NS050375].

Conflict of interest statement. None declared.

REFERENCES

1. NCBI Resource Coordinators. (2013) Database resources of the national center for biotechnology information. *Nucleic Acids Res*, **41**, D8–D20.
2. Flicek,P., Ahmed,I., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S. *et al.* (2013) Ensembl 2013. *Nucleic Acids Res*, **41**, D48–D55.
3. Pontius,J.U., Wagner,L. and Schuler,G.D. (2003) UniGene: a unified view of the transcriptome. *The NCBI Handbook*. National Center for Biotechnology Information, Bethesda, MD.
4. Matys,V., Fricke,E., Geffers,R., Gossling,E., Haubrock,M., Hehl,R., Hornischer,K., Karas,D., Kel,A.E., Kel-Margoulis,O.V. *et al.* (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, **31**, 374–378.
5. Pruitt,K.D., Tatusova,T., Brown,G.R. and Maglott,D.R. (2012) NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*, **40**, D130–D135.
6. Willis,R.C. and Hogue,C.W. (2006) Searching, viewing, and visualizing data in the biomolecular interaction network database (BIND). *Curr. Protoc. Bioinformatics*, Chapter 8, Unit 8 9.
7. Chatr-Aryamontri,A., Breitkreutz,B.J., Heinicke,S., Boucher,L., Winter,A., Stark,C., Nixon,J., Ramage,L., Kolas,N., O'Donnell,L. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res*, **41**, D816–D823.
8. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database–2009 update. *Nucleic Acids Res*, **37**, D767–D772.
9. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*, **40**, D857–D861.
10. The UniProt Consortium. (2013) Update on activities at the universal protein resource (UniProt) in 2013. *Nucleic Acids Res*, **41**, D43–D47.
11. Quevillon,E., Silventoinen,V., Pillai,S., Harte,N., Mulder,N., Apweiler,R. and Lopez,R. (2005) InterProScan: protein domains identifier. *Nucleic Acids Res*, **33**, W116–W120.
12. Tanabe,M. and Kanehisa,M. (2012) Using the KEGG database resource. *Curr. Protoc. Bioinformatics*, Chapter 1, Unit1 12.
13. Croft,D., O'Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*, **39**, D691–D697.
14. Schaefer,C.F., Anthony,K., Krupa,S., Buchoff,J., Day,M., Hannay,T. and Buetow,K.H. (2009) PID: the pathway interaction database. *Nucleic Acids Res*, **37**, D674–D679.
15. Franceschini,A., Szklarczyk,D., Frankild,S., Kuhn,M., Simonovic,M., Roth,A., Lin,J., Minguez,P., Bork,P., von Mering,C. *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*, **41**, D808–D815.
16. Choi,C., Krull,M., Kel,A., Kel-Margoulis,O., Pistor,S., Potapov,A., Voss,N. and Wingender,E. (2004) TRANSPATH–a high quality database focused on signal transduction. *Comp. Funct. Genomics*, **5**, 163–168.
17. Cerami,E.G., Gross,B.E., Demir,E., Rodchenkov,I., Babur,O., Anwar,N., Schultz,N., Bader,G.D. and Sander,C. (2011) Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res*, **39**, D685–D690.
18. Schriml,L.M., Arze,C., Nadendla,S., Chang,Y.W., Mazaitis,M., Felix,V., Feng,G. and Kibbe,W.A. (2012) Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Res*, **40**, D940–D946.
19. Basu,S.N., Kollu,R. and Banerjee-Basu,S. (2009) AutDB: a gene reference resource for autism research. *Nucleic Acids Res*, **37**, D832–D836.
20. Jia,P., Sun,J., Guo,A.Y. and Zhao,Z. (2010) SZGR: a comprehensive schizophrenia gene resource. *Mol. Psychiatry*, **15**, 453–462.
21. Kohler,S., Doelken,S.C., Rath,A., Ayme,S. and Robinson,P.N. (2012) Ontological phenotype standards for neurogenetics. *Hum. Mutat.*, **33**, 1333–1339.
22. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.
23. Iafrate,A.J., Feuk,L., Rivera,M.N., Listewnik,M.L., Donahoe,P.K., Qi,Y., Scherer,S.W. and Lee,C. (2004) Detection of large-scale variation in the human genome. *Nat. Genet*, **36**, 949–951.
24. Stenson,P.D., Mort,M., Ball,E.V., Shaw,K., Phillips,A.D. and Cooper,D.N. (2013) The human gene mutation database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet* (epub ahead of print.).
25. Konneker,T., Barnes,T., Furberg,H., Losh,M., Bulik,C.M. and Sullivan,P.F. (2008) A searchable database of genetic evidence for psychiatric disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 671–675.
26. Rzhetsky,A., Iossifov,I., Koike,T., Krauthammer,M., Kra,P., Morris,M., Yu,H., Duboue,P.A., Weng,W., Wilbur,W.J. *et al.* (2004) GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *J. Biomed. Inform.*, **37**, 43–53.
27. Davis,A.P., Wiegers,T.C., Johnson,R.J., Lay,J.M., Lennon-Hopkins,K., Saraceni-Richards,C., Sciaky,D., Murphy,C.G. and Mattingly,C.J. (2013) Text mining effectively scores and ranks the literature for improving chemical-gene-disease curation at the comparative toxicogenomics database. *PLoS One*, **8**, e58201.
28. Gotea,V., Visel,A., Westlund,J.M., Nobrega,M.A., Pennacchio,L.A. and Ovcharenko,I. (2010) Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res*, **20**, 565–577.
29. Visel,A., Minovitsky,S., Dubchak,I. and Pennacchio,L.A. (2007) VISTA enhancer browser—a database of tissue-specific human enhancers. *Nucleic Acids Res*, **35**, D88–D92.
30. Xie,B., Agam,G., Sulakhe,D., Maltsev,N., Chitturi,B. and Gilliam,T. (2012) Prediction of candidate genes for neuropsychiatric disorders using feature-based enrichment. *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, p.564–566.
31. Saad,Y. (1992) Analysis of some Krylov subspace approximations to the matrix exponential operator. *SIAM J. Num. Anal. (SINUM)*, **29**, 209–228.
32. Page,L., Brin,S., Motwani,R. and Winograd,T. (1999) The pagerank citation ranking: bringing order to the web. Stanford InfoLab.

33. Chen,J., Aronow,B.J. and Jegga,A.G. (2009) Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics*, **10**, 73.
34. Nitsch,D., Tranchevent,L.C., Goncalves,J.P., Vogt,J.K., Madeira,S.C. and Moreau,Y. (2011) PINTA: a web server for network-based gene prioritization from expression data. *Nucleic Acids Res.*, **39**, W334–W338.
35. Nitsch,D., Goncalves,J.P., Ojeda,F., de Moor,B. and Moreau,Y. (2010) Candidate gene prioritization by network analysis of differential expression using machine learning approaches. *BMC Bioinformatics*, **11**, 460.
36. Börnigen,D., Tranchevent,L.C., Bonachela-Capdevila,F., Devriendt,K., De Moor,B., De Causmaecker,P. and Moreau,Y. (2012) An unbiased evaluation of gene prioritization tools. *Bioinformatics*, **28**, 3081–3088.
37. Mulligan,C.K. and Trauner,D.A. (2013) Incidence and behavioral correlates of epileptiform abnormalities in autism spectrum disorders. *J. Autism Dev. Disord.* (epub ahead of print.).
38. Robinson,S.J. (2012) Childhood epilepsy and autism spectrum disorders: psychiatric problems, phenotypic expression, and anticonvulsants. *Neuropsychol. Rev*, **22**, 271–279.
39. Kantojarvi,K., Kotala,I., Rehnstrom,K., Ylisaukko-Oja,T., Vanhala,R., von Wendt,T.N., von Wendt,L. and Jarvela,I. (2011) Fine mapping of Xq11.1-q21.33 and mutation screening of RPS6KA6, ZNF711, ACSL4, DLG3, and IL1RAPL2 for autism spectrum disorders (ASD). *Autism Res*, **4**, 228–233.
40. Chang,S.C., Pauls,D.L., Lange,C., Sasanfar,R. and Santangelo,S.L. (2011) Common genetic variation in the GAD1 gene and the entire family of DLX homeobox genes and autism spectrum disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **156**, 233–239.
41. Vinci,G., Chantot-Bastaraud,S., El Houate,B., Lortat-Jacob,S., Brauner,R. and McElreavey,K. (2007) Association of deletion 9p, 46,XY gonadal dysgenesis and autistic spectrum disorder. *Mol. Hum. Reprod.*, **13**, 685–689.
42. Tavassoli,T., Auyeung,B., Murphy,L.C., Baron-Cohen,S. and Chakrabarti,B. (2012) Variation in the autism candidate gene GABRB3 modulates tactile sensitivity in typically developing children. *Mol. Autism*, **3**, 6.
43. Yadav,R., Gupta,S.C., Hillman,B.G., Bhatt,J.M., Stairs,D.J. and Dravid,S.M. (2012) Deletion of glutamate delta-1 receptor in mouse leads to aberrant emotional and social behaviors. *PLoS One*, **7**, e32969.
44. Johnson,R., Hoeller,J., Arendsen,A., Risberg,T. and Kopylenko,D. (2005) *Professional Java Development with the Spring Framework. (Recommendation)*. Wrox Press Ltd., Birmingham, UK.
45. Potociar,M. (2009) JSR 311: JAX-RS: the java API for RESTful web services. *Technical report*. Oracle.
46. Kawaguchi,K., Vajjhala,S. and Fialli,J. (eds), (2006) *The JavaTM Architecture for XML Binding (JAXB) 2.1*. Sun Microsystems, Inc.
47. Madduri,R., Dave,P., Sulakhe,D., Lacinski,L., Liu,B., and Foster,I. (2013). Experiences in building a next-generation sequencing analysis service using galaxy, globus online and Amazon web service. In: *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery* (XSEDE '13). ACM, New York NY, USA, Article 34, 3 pages.
48. Lukashin,I., Novichkov,P., Boffelli,D., Paciorkowski,A.R., Minovitsky,S., Yang,S. and Dubchak,I. (2011) VISTA region viewer (RViewer)—a computational system for prioritizing genomic intervals for biomedical studies. *Bioinformatics*, **27**, 2595–2597.